

# Model Independent Search in 2-Dimensional Mass Space

G. Anagnostou

*I. Physikalisches Institut B, RWTH Aachen, Germany*

## Abstract.

A model independent method to search for particles of unknown masses in events with missing energy is presented. The only assumption is the topology of the decay chain. The method is tested in events with top pairs decaying leptonically with the presence of two neutrinos in the final state. Possible applications are searches for a heavy resonance decaying to top pairs as well as next generation heavy quarks. Similar topologies are predicted by supersymmetric models with R-parity conservation resulting in final states with two invisible neutralinos.

**Keywords:** searches

## INTRODUCTION

Could we find the mass of the top quark and W boson in LHC from top-pairs decaying leptonically in the hypothetical case in which both  $m_t$  and  $m_W$  were unknown? Could we establish a discovery by observing the mass peaks above background for both particles without any assumption about the underlying theory?

Standard Model has predicted the mass of the W boson which was later measured accurately by LEP. Top mass has also been measured in both Tevatron and LHC. The reason that such a question is interesting is the similarity with BSM theory models which predict new invisible particles (e.g neutralinos). It is believed by many physicists that such theories might be the next step in our understanding of the physical world as they provide (among others) an explanation for the nature of dark matter. For example, supersymmetric models with R-Parity conservation predict event topologies with at least two neutralinos. The existence of two invisible particles in the decay chain makes the extraction of the unpredicted by the model masses a difficult problem [1], [2], [3]. Similar topologies might exist in the decays of a hypothetical heavier quark from a fourth generation or a heavy resonance decaying to top pairs.

Observation of an excess of events with large missing energy would provide evidence for the existence of new physics but could not help further in our understanding of what the new physics is. Several  $M_{ET}$ -like observables have been proposed for this purpose ( $M_{ET}$ ,  $H_T$ ,  $M_{eff}$  etc). A common characteristic of all of them is that new physics would appear as an excess of events in the tail of a distribution. The establishment of a discovery would be a difficult problem due to the small difference in the shape between signal and background. In this case, a very good understanding of the tail would be necessary, a challenging task at least at an early stage. But let's say

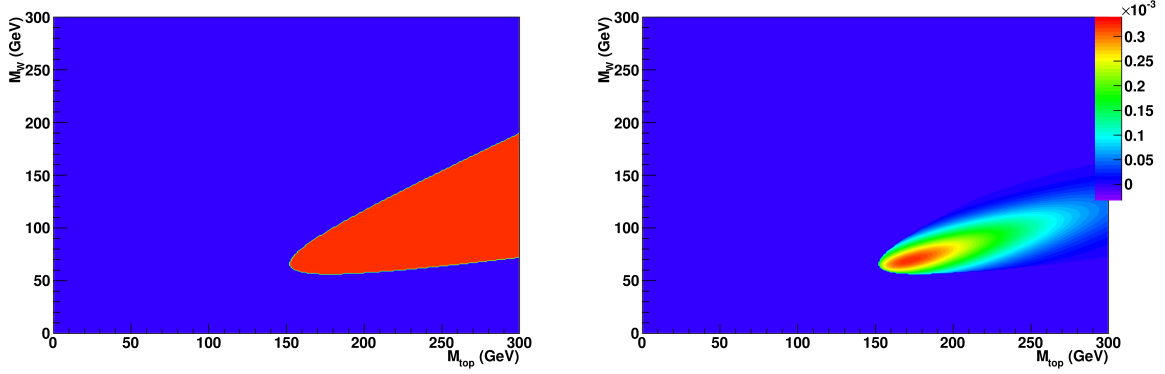
that LHC experiments observe an excess of events in the tail of a  $M_{ET}$ -like observable and have confidence in the result. What can we say about new physics except that it exist?

Understanding the physics is much easier in possible BSM signals which do not predict any invisible particles. In this case, the invariant masses can be reconstructed up to combinatoric ambiguities. Knowledge of the masses provides full understanding of the event kinematics and subsequently allows boosts to the correct rest frames of the decaying particles. Angular distributions in these rest frames provide the best observables to discriminate between different spin hypothesis. Knowledge of both mass and spin of the new particles is a very important step towards understanding what the new physics is.

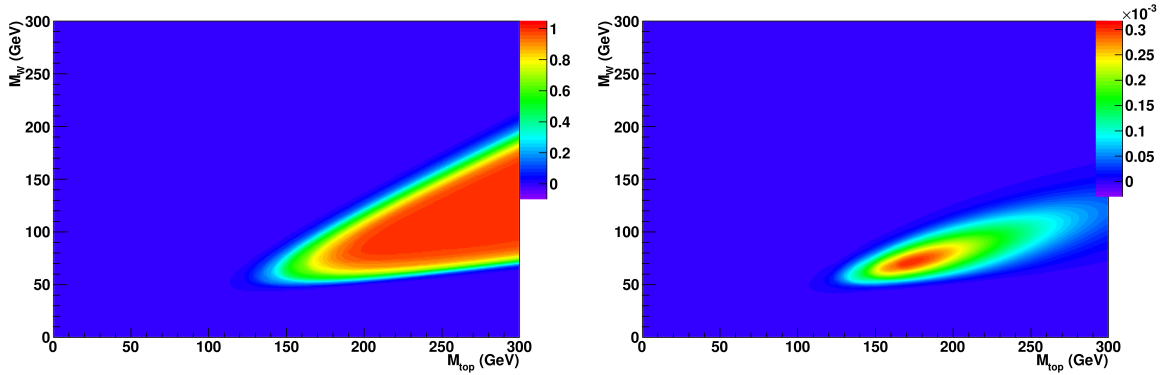
In addition, signal events are contained in a small region of mass space in contrast to the higher in dimensions kinematic phase space in which selection cuts are often applied. For example, in the supersymmetric pseudo-scalar MSSM Higgs decay  $A \rightarrow Zh, Z \rightarrow l^+l^-, h \rightarrow b\bar{b}$  the reconstruction of both A, h hypothetical Higgs bosons is possible. The new physics events would then be observed as a peak in the 2-dimensional  $m_A, m_h$  mass plane [4]. In this case, there is significant difference in shape between signal and background. The establishment of a discovery is an easier task as a data-driven estimation of the background can be performed from the sidebands.

Searching for resonances in mass-space is model independent as the signal extraction does not need any change that depends on the hypothetical model. The selection applied can be simple energy thresholds to ensure correct reconstruction as well as identification of the final state objects.

It is often believed that reconstruction of particles masses in topologies with two invisible particles is not feasible. In the rest of this study a counter example is presented with the simultaneous reconstruction of W



**FIGURE 1.** (Left): Solvability for the correct solution of a single top-pair event in the 2-dimensional mass plane. (Right): Solvability weighted with the PDFs and normalized to unit volume for the same solution and event.



**FIGURE 2.** (Left): Solvability (for the correct solution) of 300 test events produced from the same single event by smearing of the momenta of leptons, jets and transverse missing energy. (Right): Average weighted solvability for the same test events normalized to unit volume.

boson and top quark in the dilepton top-pair final state in a p-p collider like LHC. The method can also be used to search in a model independent way for a new generation  $t'$  and any resonance decaying to top pairs  $pp \rightarrow X' \rightarrow t\bar{t}$ .

## EVENT SIMULATION AND SELECTION

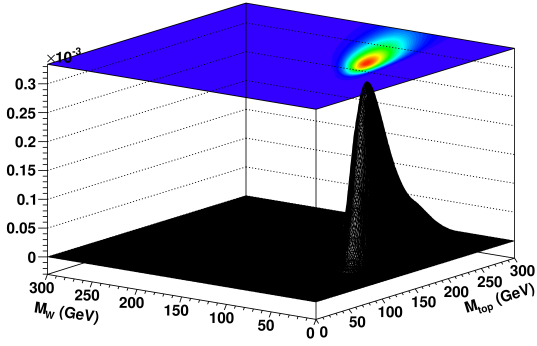
All Monte Carlo sample used in this study were generated using Pythia 6 [5] except the production of a next generation quark via  $pp \rightarrow f\bar{f}, q\bar{q} \rightarrow t'\bar{t}'$  for which Pythia 8 was used [6]. Top pair events with  $m_t = 172.5$  GeV and  $m_W = 80.4$  GeV were generated corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$ . For the same integrated luminosity samples of Drell-Yan ( $Z \rightarrow e^+e^-, \mu^+\mu^-$ ) and diboson backgrounds (WW, WZ, ZZ) were also created.

The performance of the method to new resonances decaying to top pairs was tested using a sample of 300  $pp \rightarrow Z' \rightarrow t\bar{t}$  events, with the new gauge boson generated with  $m_{Z'} = 500$  GeV and  $\Gamma_{Z'} = 15$  GeV. Finally, for

the search for a new generation quark, a sample of 300  $pp \rightarrow f\bar{f}, q\bar{q} \rightarrow t'\bar{t}'$  events was produced with  $m_t = 250$  GeV. Pythia cross-sections were used in all cases except top pairs for which a more accurate calculation was used [7].

### Jet reconstruction

All stable particles except neutrinos were used as an input to Fastjet-2.4.2 in order to cluster particles into generated jets [8]. Jet reconstruction was performed with anti- $K_T$  algorithm using a cone of 0.5. The generated jets were then smeared according to the jet reconstruction performance of a typical LHC detector [11]. The ratio of the reconstructed to the partonic jet energy was used to correct the jet energy scale. This was performed in 3  $P_T$  bins ( $E_T < 50$  GeV,  $50 \text{ GeV} \leq E_T < 150$  GeV,  $E_T \geq 150$  GeV) and 2 pseudorapidity bins ( $|\eta| < 1.5$ ,  $1.5 < |\eta| < 2.5$ ). The correction factors were calculated from the top-



**FIGURE 3.** Average PDF weighted Solvability for the 300 test events produced from the same initial top pair event. The distribution is normalized to unit volume.

pair sample using the bjets matched with a bquark. All jets matched with a bquark were given 50% chance to be tagged as bjets.

#### *Leptons and Missing Energy*

Missing energy was calculated by summing vectorially all reconstructed jets, muons and electrons and then reversing the sign. Both jets and  $M_{ET}$  resolutions were checked to be as described in [11]. Muon and electron energies were smeared at the percent level which is a typical resolution for an LHC detector.

#### *Event Selection*

Events were selected as top pair candidates by requiring all objects of the final state to be above an energetic threshold in order to be well reconstructed: at least two leptons (electrons or muons) with  $P_T \geq 20$  GeV, and at least two bjets with  $P_T \geq 20$  GeV. In case the event had more than two bjets or leptons, the two most energetic ones were selected. All jets and leptons were required to be in the pseudorapidity region  $|\eta| < 2.5$ . Finally, events with transverse missing energy less than 20 GeV were rejected. Events satisfying the above selection were used as input to the algorithm described in the next section.

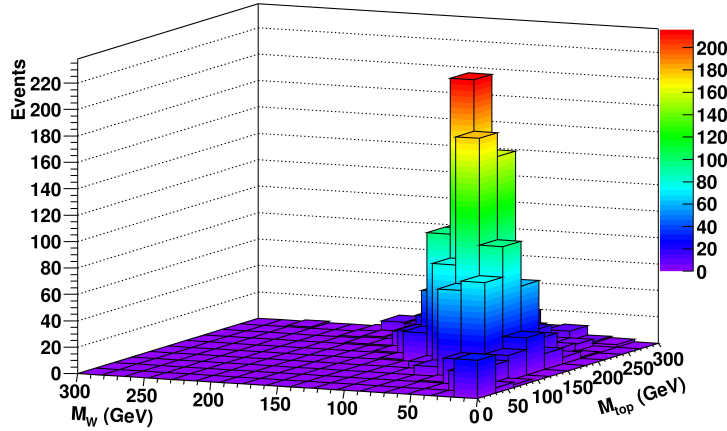
## THE METHOD

The kinematics of  $t\bar{t}$  dilepton events can be expressed by two linear and six non linear equations (Appendix). The system is solvable with respect to the unknown neutrino and antineutrino four vectors, provided that the masses

of top quark and W boson are known. An analytical solution for the equations of the top-pair system is described in [12], [13]. The transverse missing energy components, the momentum of bquarks and leptons as well as the masses of top quark and W boson are used as an input for the analytical solution algorithm. Each possible input can give 0, 2 or four different solutions for the unknown neutrino and antineutrino momentum components. In addition, there are two possible combinations of bjets and leptons that could originate from the same top quark, giving in total up to eight solutions. So, given the momenta of leptons and bjets together with  $M_{ET_x}, M_{ET_y}$  there can be up to eight possible neutrino and antineutrino momentum vectors for a given  $m_t$  and  $m_W$ . Knowledge of the momenta of the invisible neutrinos allows full kinematic reconstruction of the event including the four-vectors of W-bosons, top quarks and the  $t\bar{t}$  system.

The initial goal was to "rediscover" W boson and top quark by looking at LHC data without any assumption about the underlying theory, except the topology of top-pairs. As the masses of the particles are unknown the only option is to test every point of the  $m_t, m_W$  plane for possible solutions. The mass plane can be scanned in steps of 1 GeV to produce the area in which each one of the solutions exists or not. Solvability for a single event can be defined as the existence (or not) of a specific solution in a specific mass point. An example of such a solution area for a single top pair event is plotted in Figure 1 (left). By observing several signal events, it can be seen that solvability is bounded only from below for both  $m_t$  and  $m_W$ .

Due to the finite collision energy there is also an upper limit to the allowed masses produced. The center of mass energy of the partons participating in the hard scattering has to be smaller than the LHC collision energy. For a p-p collider this energy limit can be expressed fully by the parton distribution functions (PDFs) of the proton in the following way: each solution allows full reconstruction of event kinematics, including the estimation of the energy E and  $P_Z$  momenta component of the  $t\bar{t}$  system. These variables can be easily transformed to the fraction of beam energy of the two partons participating in the hard scattering ( $x_{1,2} = E \pm p_z/2$ ). So each parton with fraction  $x_i$  can be assigned with a probability  $F(x_i)$  to originate from a proton-proton collision. By multiplying the probabilities of the two partons a weight per mass point can be assigned for each solution. As there are more than one possible leading order parton-parton interactions ( $u\bar{u}, \bar{u}u, d\bar{d}, \bar{d}d, gg$ ) weights from all possible combinations are summed to estimate a final event weight per solution and mass point. The weight can be written as  $\sum F(x_i)F(\bar{x}_i)$ , where the summation is over the parton combinations and F is the CTEQ6.1 PDF set with momentum transfer  $Q^2 = m_t^2$  [9]. For the estimation of the PDF values the LHAPDF interface was used [10].



**FIGURE 4.** Reconstructed  $m_t$  and  $m_W$  per event for a sample of signal and background events corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$ .

The PDF weight normalized to unit volume provides an upper bound for the mass values of both  $m_t$  and  $m_W$ . An example of solvability weighted by the PDFs for the same single top pair event is plotted in Figure 1 (right). It is interesting to mention that the preferred mass point is not the one with the lowest  $m_t$  and  $m_W$  values as one might have guessed from the fact that PDFs favour lower mass values. Use of the solvability together with a matrix element weight which depends on the model has been proposed for top mass estimation in a single mass dimension [14]. This proposal has evolved to the matrix element weighting method in Tevatron [15]. What is proposed in this study is a general multi-dimensional search method in mass space using only model independent PDF weights rather than a top mass measurement method using matrix elements.

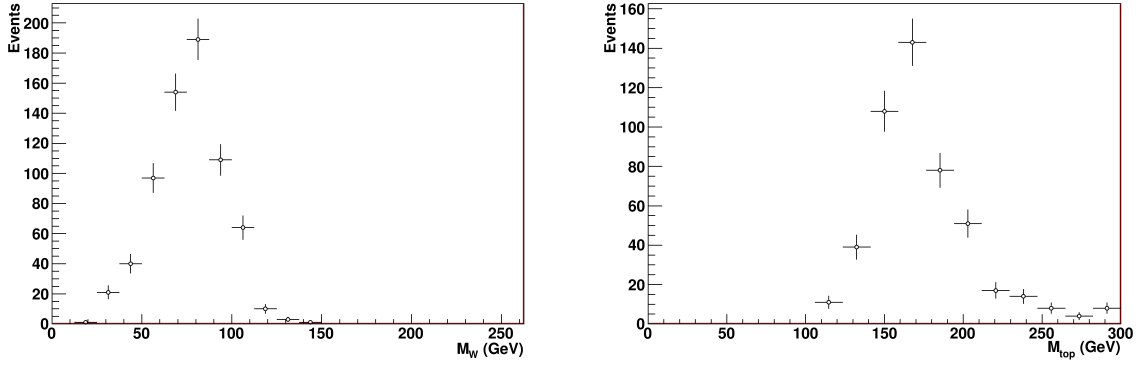
Detector effects can change the momenta of the leptons and jets making a solvable event not-solvable. In many cases solvability can be recovered by smearing the leptons and jets according to detector resolution. For each initial event,  $N$  test events can be created by smearing the leptons, jets and missing energy components of the recorded event according to the detector resolutions. For these test events, solvability can be defined as the fraction of them for which a specific solution exist. Solvability of the 300 test events created by the initial single top pair event is plotted in Figure 2 (left).

For each test event, solvability of a solution (which is either zero or one) can be multiplied with the PDF weight calculated for the specific solution and mass point. The value obtained is averaged over all 300 test events and normalized to unit volume. Such a distribution can be constructed for all possible solutions. An example is given in Figure 2 (right) for the correct solution of the ini-

tial top pair event. Among all solutions, the one with the highest PDF weight is chosen. The final  $m_t$  and  $m_W$  estimation for this event is the mass point where the distribution of the preferred solution is maximized (Figure 3). The above procedure gives a single mass point per event. The other option is to construct a complicated likelihood in order to exploit all the available information. A single mass entry is preferred so as to reconstruct invariant mass distributions in which robust discovery techniques can be applied. It is worth emphasizing the power of the PDF weights to choose a preferred solution. Not all solutions are as likely to originate from a proton-proton collision and the parton distribution functions can distinguish one of them. This might be applicable to other cases with combinatoric backgrounds such as reconstruction of chains with visible particles.

By applying the method in signal and background events corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$ , the 2-dimensional mass distribution can be created (Figure 4). In order to produce the invariant mass distribution for the lighter particle, all events around the heavier one can be selected, according to the mass resolution of the detector. For W boson, all events with  $150 \text{ GeV} < m_t < 190 \text{ GeV}$  in the 2-dimensional mass distribution can be selected (Figure 5, left). In a similar way for top quark, all events with  $70 \text{ GeV} < m_W < 90 \text{ GeV}$  are selected (Figure 5, right). The 2-dimensional mass distribution as well as the one-dimensional ones show that the method works: both top quark and W boson would be observed without any a priori knowledge of their masses or of the underlying theory, in a model independent way. The method described can be tested using LHC data. If it works for top pairs, it is likely to work for any of the searches described in the next section.





**FIGURE 5.** Reconstructed  $m_W$  when requiring  $m_t$  to be in the range 150-190 GeV (left) as well as reconstructed  $m_t$  when requiring  $m_W$  to be in the range 70-90 GeV (right), for a sample of signal and background events corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$ .

## APPLICATIONS

### Search for a resonance decaying to top pairs

An interesting application is the search for a heavy particle decaying to top pairs  $pp \rightarrow X' \rightarrow t\bar{t}$  in the dilepton final state. The later has less background than the semileptonic and hadronic top-pair final states. In addition, it allows easier boosted top reconstruction from its final state objects as top is reconstructed from a single bjet and a lepton rather than 3 closely spaced jets. In order to test the method, a sample of 300 events of  $pp \rightarrow Z' \rightarrow t\bar{t}$  were created, with  $m_{Z'}=500 \text{ GeV}$  and  $\Gamma_{Z'}=15 \text{ GeV}$ .

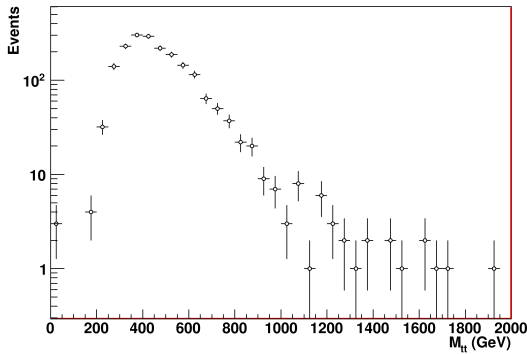
The average weighted solvability of the test events is used to get an estimated  $m_t$  and  $m_W$  per event, as described in the previous section. The solution and mass point with the highest PDF weight is chosen. A full kinematic

reconstruction can now take place for each of the test events: Starting from the neutrino momenta, the four-vectors of W boson and top quark can be calculated for both branches. Knowledge of the above allows the reconstruction of the  $t\bar{t}$  system and subsequently its invariant mass. The average  $m_{t\bar{t}}$  of all solvable test events is used as an estimate of the mass of a possible  $X' \rightarrow t\bar{t}$  decay (Figure 6). Such a resonance would be observed as a peak in the  $m_{t\bar{t}}$  distribution, provided that its width is not very wide. The 2-dimensional reconstructed  $m_t$  and  $m_W$  per event can be seen in Figure 7 (left). The reconstructed  $m_{Z'}$  for the same events is shown in Figure 7 (right). The invariant mass distribution is an interesting observable to look for a heavy new particle as there are very few Standard Model events in the high mass range.

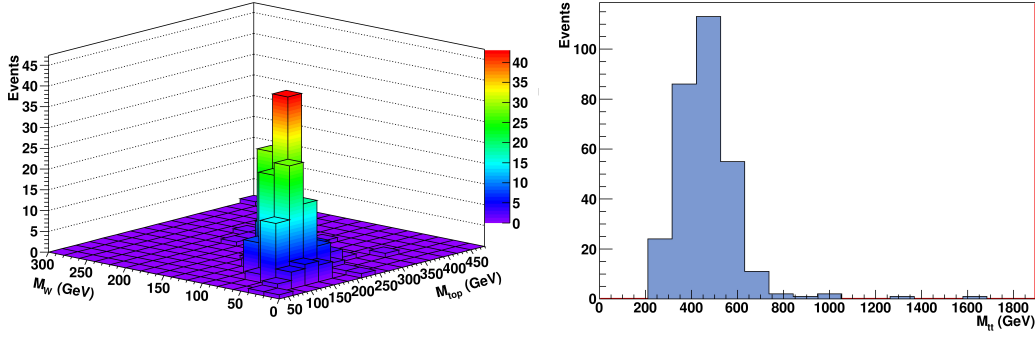
### Search for a new generation heavy quark

The 2-dimensional mass reconstruction described is model independent and would work for any particle decaying like top quark or W boson. So one could look for anything decaying as  $YY \rightarrow XbXb \rightarrow llbb\nu_l\nu_l$  ( $Y \rightarrow Xb$ ,  $X \rightarrow l\nu_l$ ). The two chains can be identical or a particle anti-particle pair as in both cases the masses are the same.

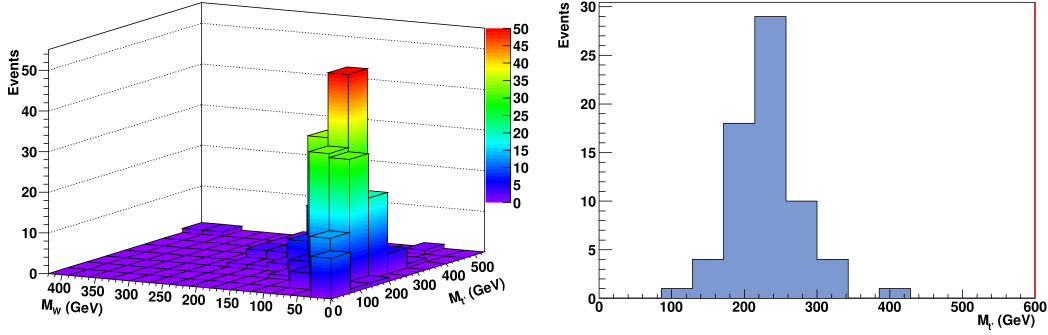
In case Y is a next generation heavy quark  $t'$  the topology would be  $t'\bar{t}' \rightarrow W^+bW^-\bar{b} \rightarrow l^+l^-b\bar{b}\nu_l\bar{\nu}_l$ , ( $t' \rightarrow Wb$ ). Another possibility is to search for X particle instead of Y, where X could be a charged higgs boson. To demonstrate the reconstruction of a  $t'\bar{t}'$  signal, a sample of 300 events were used with  $m_{t'}=250 \text{ GeV}$ . The 2-dimensional mass reconstruction can be seen in Figure 8 (left). The one-dimensional  $t'$  mass reconstruction can be seen in Figure 8 (right).



**FIGURE 6.** Reconstructed  $m_{t\bar{t}}$  for a sample of top pair and background events corresponding to an integrated luminosity of  $1 \text{ fb}^{-1}$ .



**FIGURE 7.** Reconstructed  $m_t$  and  $m_W$  per event for a sample of 300  $Z' \rightarrow t\bar{t}$  generated with  $m_{Z'} = 500$  GeV (left) as well as reconstructed  $m_{Z'}$  for the same events (right).



**FIGURE 8.** Reconstructed  $m_{t'}$  and  $m_W$  per event for a sample of 300  $t'\bar{t}' \rightarrow W^+ b W^- \bar{b} \rightarrow l^+ l^- b \bar{b} \nu_l \bar{\nu}_l$  events generated with  $m_{t'} = 250$  GeV (left) as well as reconstructed  $m_{t'}$  when requiring  $m_W$  to be in the range 70-90 GeV (right).

## Top pair identification

Another possible application of the 2-dimensional mass reconstruction is the identification of top pair events. For many searches performed in final states with missing energy top pairs are the most significant SM background. The  $M_{ET}$ -like observables used to establish a discovery have their tail populated by top pair events. So, by identifying them using mass constraints in the 2-dimensional  $m_t$  and  $m_W$  plane, we can suppress significantly the most important source of a possible fake discovery.

## Top mass measurement

The 2-dimensional mass reconstruction can also be used for an improved top mass measurement. The simultaneous reconstruction of  $m_W$  together with  $m_t$  gives a handle to control the main systematic effect of the measurement: the jet energy scale. Calibration of this scale could be based on  $m_W$  reducing significantly the effect on the reconstruction of  $m_t$ .

## CONCLUSIONS

Mass space is the natural space to search for new particles. Observation of mass peaks above background allows robust discovery using data-driven background estimation from the sidebands. But most important, reconstruction of the unknown masses gives valuable insights to what the new physics is. The search is model independent as the only assumption is the decay topology. As a proof of principle, mass peaks of both top quark and W boson can be produced from leptonic top pair decays from simulated events as well as using LHC data. It is shown that the method can be used to reconstruct mass peaks from heavy resonances decaying to top pairs as well as next generation heavy quarks decaying with the same event topology. Possible applications to supersymmetric mass reconstruction will be described in a future publication. Other applications include top pair identification in order to reject them from the tail of  $M_{ET}$ -like observables for supersymmetric searches, as well as use of the  $m_W$  to control the jet energy scale for an improved top mass measurement.

## ACKNOWLEDGMENTS

I would like to thank Lars Sonnenschein for providing the software used in [13]. Many thanks to Andrei Ostapchouk and Bob McElrath for discussions and suggestions concerning the PDF weights.

## APPENDIX

The top pair event topology has the following constraints:

$$M_{ETx} = p_{vx} + p_{\bar{v}x} \quad E_v^2 = p_{vx}^2 + p_{vy}^2 + p_{vz}^2$$

$$M_{ETy} = p_{vy} + p_{\bar{v}y} \quad E_{\bar{v}}^2 = p_{\bar{v}x}^2 + p_{\bar{v}y}^2 + p_{\bar{v}z}^2$$

$$m_{W^+}^2 = (E_{l^+} + E_v)^2 - (p_{lx^+} + p_{vx})^2 - (p_{ly^+} + p_{vy})^2 - (p_{lz^+} + p_{vz})^2$$

$$m_{W^-}^2 = (E_{l^-} + E_{\bar{v}})^2 - (p_{lx^-} + p_{\bar{v}x})^2 - (p_{ly^-} + p_{\bar{v}y})^2 - (p_{lz^-} + p_{\bar{v}z})^2$$

$$m_t^2 = (E_b + E_{l^+} + E_v)^2 - (p_{bx} + p_{lx^+} + p_{vx})^2 - (p_{by} + p_{ly^+} + p_{vy})^2 - (p_{bz} + p_{lz^+} + p_{vz})^2$$

$$m_{\bar{t}}^2 = (E_{\bar{b}} + E_{l^-} + E_{\bar{v}})^2 - (p_{\bar{b}x} + p_{lx^-} + p_{\bar{v}x})^2 - (p_{\bar{b}y} + p_{ly^-} + p_{\bar{v}y})^2 - (p_{\bar{b}z} + p_{lz^-} + p_{\bar{v}z})^2$$

## REFERENCES

1. Hsin-Chia Cheng, John F. Gunion, Zhenyu Han, Guido Marandella, Bob McElrath, (UC, Davis), Mass determination in SUSY-like events with missing energy, JHEP 0712:076 2007, arXiv:0707.0030.
2. Hsin-Chia Cheng, (UC, Davis), Dalit Engelhardt, (Boston U.), John F. Gunion, Zhenyu Han, (UC, Davis), Bob McElrath, (CERN), Accurate Mass Determinations in Decay Chains with Missing Energy, Phys.Rev.Lett.100:252001 2008, arXiv:0802.4290.
3. Hsin-Chia Cheng, John F. Gunion, Zhenyu Han, (UC, Davis), Bob McElrath, (CERN), Accurate Mass Determinations in Decay Chains with Missing Energy. II, Phys.Rev.D80:035020 2009, arXiv:0905.1344.
4. G.Anagnostou, G.Daskalakis, Search for the MSSM  $A \rightarrow Zh$  with  $Z \rightarrow l^+l^-$ ,  $h \rightarrow b\bar{b}$ , Journal of Physics G: Nuclear and Particle Physics, Volume 34, Number 5, May 2007.
5. T. Sjostrand, S. Mrenna and P. Skands, PYTHIA 6.4 Physics and Manual, JHEP05 (2006) 026.
6. T. Sjostrand, S. Mrenna and P. Skands, A Brief Introduction to PYTHIA 8.1, Comput. Phys. Comm. 178 (2008) 852 arXiv:0710.3820.
7. N.Kidonakis, Top quark cross sections and differential distributions, arXiv:1105.3481v1.
8. M. Cacciari and G.P. Salam, Dispelling the  $N^3$  myth for the Kt jet-finder, Phys. Lett. B 641 (2006) 57 hep-ph/0512210, M. Cacciari, G.P. Salam and G. Soyez, <http://fastjet.fr>
9. J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P. Nadolsky, W.K. Tung, New Generation of Parton Distributions with Uncertainties from Global QCD Analysis, arXiv:hep-ph/0201195v3.
10. M.R. Whalley, (Durham U.) , D. Bourilkov, R.C. Group, (Florida U.), The Les Houches accord PDFs (LHAPDF) and LHAGLUE, A Workshop on the Implications of HERA and LHC Physics, Hamburg, Germany, 21-24 Mar 2005, hep-ph/0508110.
11. CMS Collaboration, CMS Physics TDR: Volume I (PTDR1), Detector Performace and Software, CERN-LHCC-2006-001, 2 February 2006.
12. L. Sonnenschein, Algebraic approach to solve t anti-t dilepton equations, Phys.Rev.D72, 095020, 2005.
13. L. Sonnenschein, Analytical solution of  $t\bar{t}$  dilepton equations, Phys. Rev. D 73, 054015, 2006.
14. R. H. Dalitz and G. R. Goldstein, The decay and polarization properties of the top quark, Phys. Rev. D45 (1992) 1531. doi:10.1103/PhysRevD.45.153.
15. D0 Collaboration, Measurement of the top quark mass in final states with two leptons Phys. Rev. D80, 092006 (2009).